Journal of Software Engineering, Information and Communication Technology (SEICT)

Journal homepage: https://ejournal.upi.edu/index.php/SEICT

Application of Deep Learning using Convolutional Neural Network (CNN) Algorithm for Gesture Recognition

Ahmad Abuzar Alhamdani

School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Indonesia
Correspondence: E-mail: rua26x@gmail.com

ABSTRACT

Gesture recognition is a fascinating method of human-computer interaction that goes beyond traditional means such as keyboards, pointers, and joypads. In gesture recognition, Convolutional Neural Network (CNN) algorithms are utilized in Deep Learning to train models using datasets comprising gesture images. The training process involves pattern recognition and identification of crucial features from gesture images, followed by evaluation to measure the model's accuracy. Gesture recognition holds immense potential across various fields, including human-computer interaction, gaming, healthcare, and autonomous vehicles, and continues to be a focus of research and development in the future.

ARTICLE INFO

Article History:
Submitted/Received 25 April 2023
First Revised 01 May 2023
Accepted 23 May 2023
First Available online 26 May 2023
Publication Date 01 June 2023

Keyword:
CNN,
Deep learning,
Gesture recognition.

1. INTRODUCTION

With the advancement of technology, there are increasingly more options available for interacting with computers (Setiawan, 2018). Gesture recognition is one of the alternatives that can be used to interact with computers (Ridwang, 2018). Gestures can be applied in various applications such as command systems, robotics, gaming, sign language, and so on. Machine learning is one of the applications of artificial intelligence (Khan et al., 2012). The use of machine learning in computer vision is closely related to deep learning, where computer scientists draw inspiration from the natural world (Arifin et al., 2021).

Machine learning can be categorized into three main categories: supervised learning, unsupervised learning, and reinforcement learning (Dasgupta and Nath, 2016). In supervised learning, models are trained using labeled data, where the input data is accompanied by corresponding target labels (Yan and Wang, 2022). Unsupervised learning, on the other hand, aims to find patterns and structures in unlabeled data without predefined outputs (Ando et al., 2005). Reinforcement learning involves an agent learning to make decisions based on rewards obtained from interacting with an environment (Abdulhai et al., 2003). These three categories provide a comprehensive framework for solving a wide range of problems and have contributed to significant advancements in artificial intelligence (Roihan et al., 2020).

Deep Learning is an algorithm in machine learning that utilises artificial neural networks (ANN) as its foundation (Wahyuni and Sulaeman, 2020). Artificial neural networks are structures commonly used for classification tasks (Ju et al., 2018). In this mechanism, the object to be classified is presented to the network through the activation of artificial neurons within the input layer (Choldun and Surendo, 2018).

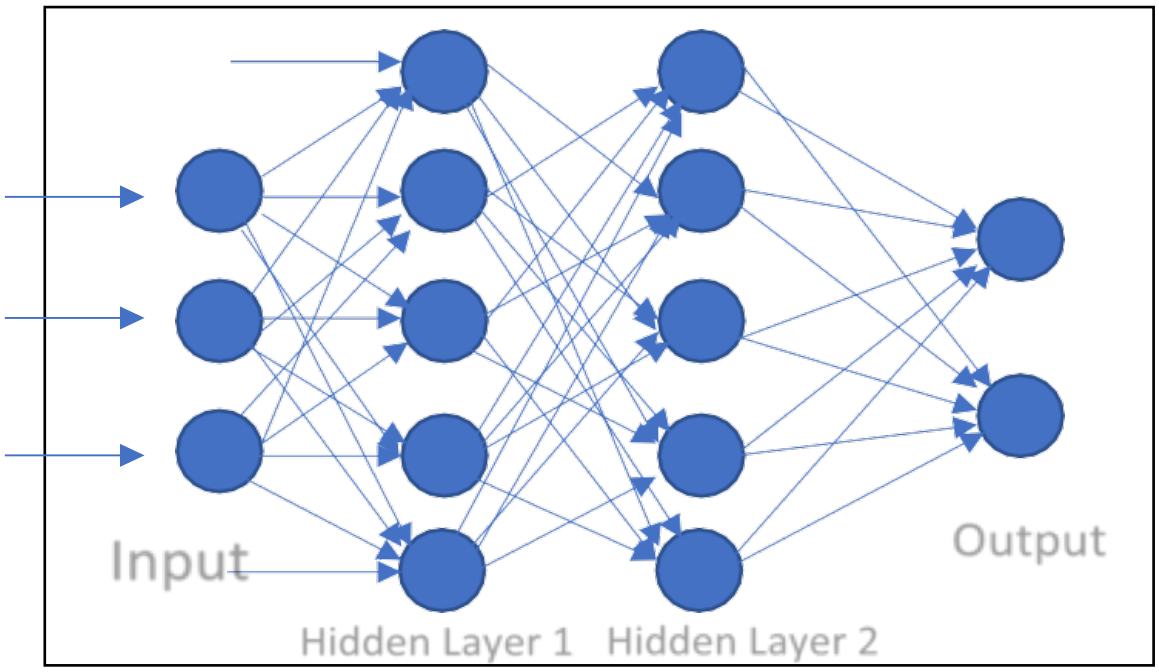

Figure 1. Artificial neural network

DOI: https://doi.org/10.17509/seict.v2i1.34673
p- ISSN 2774-1656 e- ISSN 2774-1699

Convolutional Neural Network (CNN) is widely used for image classification, object recognition, and detection tasks (Aamir et al., 2018). CNN consists of three main layers: convolution, pooling, and classification (Hu et al., 2015).

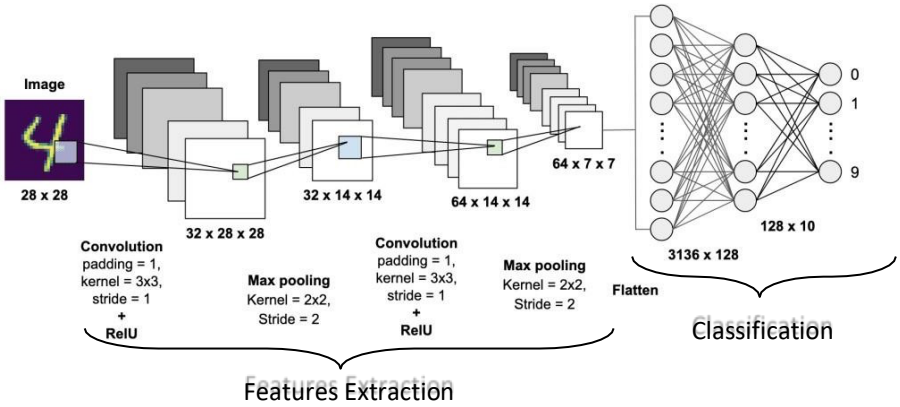

Figure 2. Convolutional neural network

By utilizing deep learning algorithms, a model will be constructed to be used in this research. The data used in this case consists of images captured using the Leap Motion device.

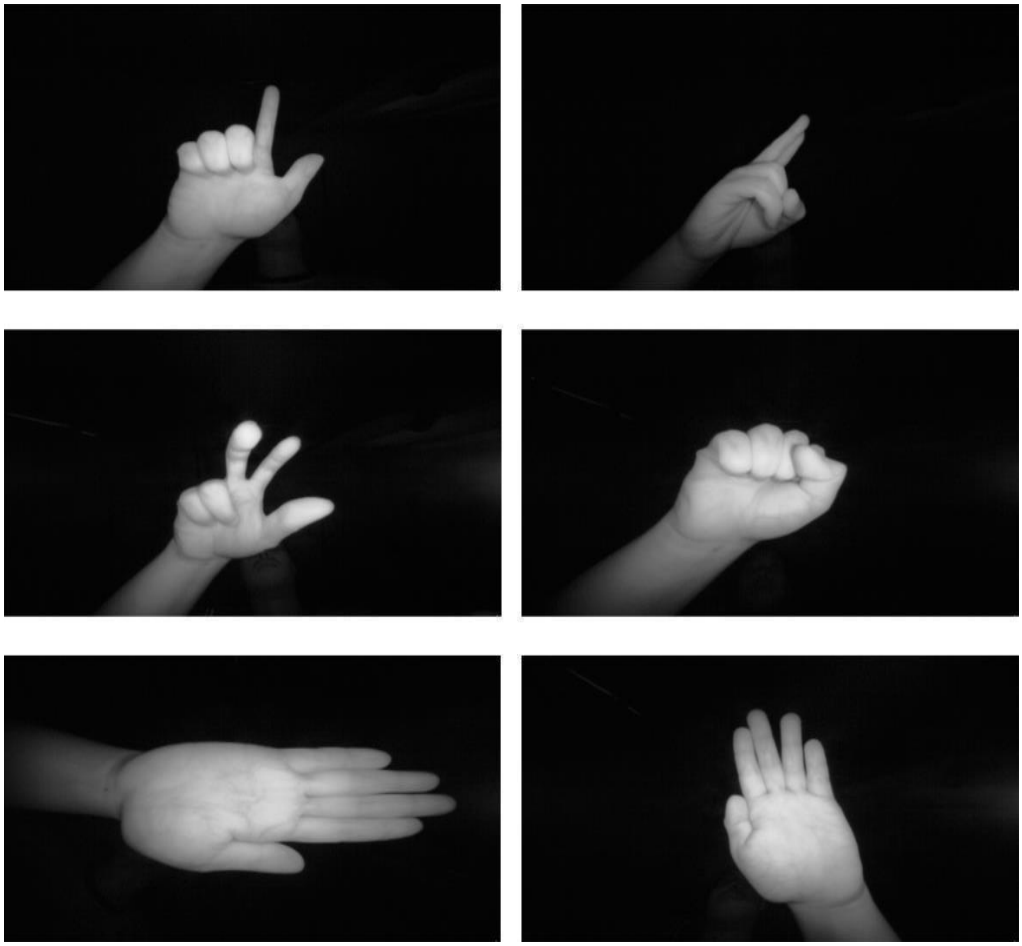

Figure 3. Gesture sample from the data set

DOI: https://doi.org/10.17509/seict.v2i1.34673
p- ISSN 2774-1656 e- ISSN 2774-1699

This research utilizes a real-time hand gesture recognition system based on OpenCV and employs the histogram of oriented gradients (HOG) and Haar Cascade classifier algorithms to classify various hand shapes (Rijanandi et al., 2023).

2. RESEARCH METHOD

The research method employed in this study is quantitative. The accuracy of training and validation is calculated using the CNN algorithm for the gesture recognition system (Kurniawan and Mustikasari, 2021).

2.1. Data gathering

Data collection for gesture recognition is conducted using the Leap Motion camera, as shown in **Figure 3**, with a resolution of 240x640 pixels.

2.2 Training Data

1. Features Extraction
   Features Extraction is performed by creating a CNN model consisting of two main parts: feature extraction and classification (Al-Doori et al., 2021). The feature extraction part includes convolutional layers and pooling layers, as shown in **Figure 2**.
2. Classification
   Classification consists of two main layers, namely the flatten layer and the dense layer, which serve as the output of the prediction model created, as shown in **Figure 2**. Subsequently, a test will be conducted (Kaliyar et al., 2021).
3. Gesture Recognized
   The gestures will be recognized after conducting a test on the model, which has been evaluated for accuracy and validation.

3. RESULTS AND DISCUSSION

The data obtained from the Leap Motion device consists of grayscale images with a resolution of 240x640 pixels, as shown in **Figure 3**. The dataset contains a total of 6000 images. In the next step, the data will be trained to obtain the desired model. During the training phase, the image data will be stored in an array and undergo features extraction using convolutional layers and pooling layers (Mesut et al., 2020). The extracted features will then be used for classification. In the classification stage or in a fully-connected layer, the desired classification results will be obtained and will be used for gesture recognition (Barbhuiya et al., 2021).

DOI: https://doi.org/10.17509/seict.v2i1.34673
p- ISSN 2774-1656 e- ISSN 2774-1699

Layer (type)	Output Shape	Param #
conv2d_30 (Conv2D)	(None, 116, 316, 32)	832
max_pooling2d_30 (MaxPooling	(None, 58, 158, 32)	0
conv2d_31 (Conv2D)	(None, 56, 156, 64)	18496
max_pooling2d_31 (MaxPooling	(None, 28, 78, 64)	0
conv2d_32 (Conv2D)	(None, 26, 76, 64)	36928
max_pooling2d_32 (MaxPooling	(None, 13, 38, 64)	0
flatten_10 (Flatten)	(None, 31616)	0
dense_20 (Dense)	(None, 128)	4046976
dense_21 (Dense)	(None, 10)	1290

Total params: 4,104,522
Trainable params: 4,104,522
Non-trainable params: 0

Figure 4. Convolutional Neural Network Model Being Used

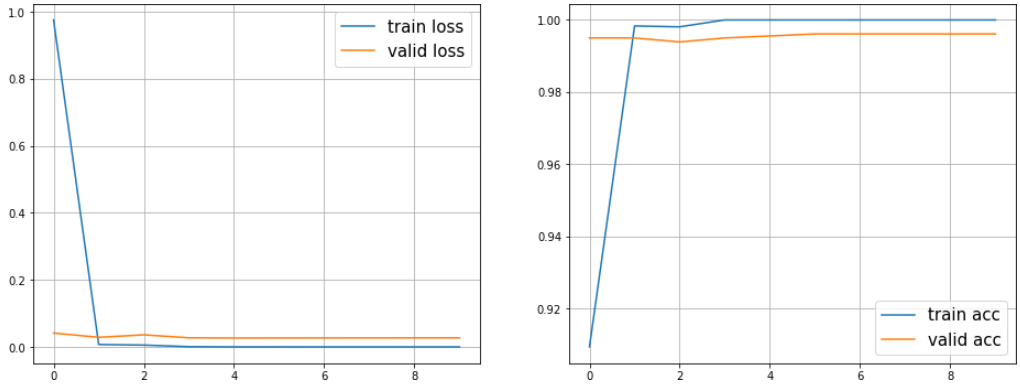

Figure 5. Accuracy and Loss when training data

During the testing phase, the results of training accuracy, training loss, validation accuracy, and validation loss are obtained, as shown in **Figure 5**.

DOI: https://doi.org/10.17509/seict.v2i1.34673
p- ISSN 2774-1656 e- ISSN 2774-1699

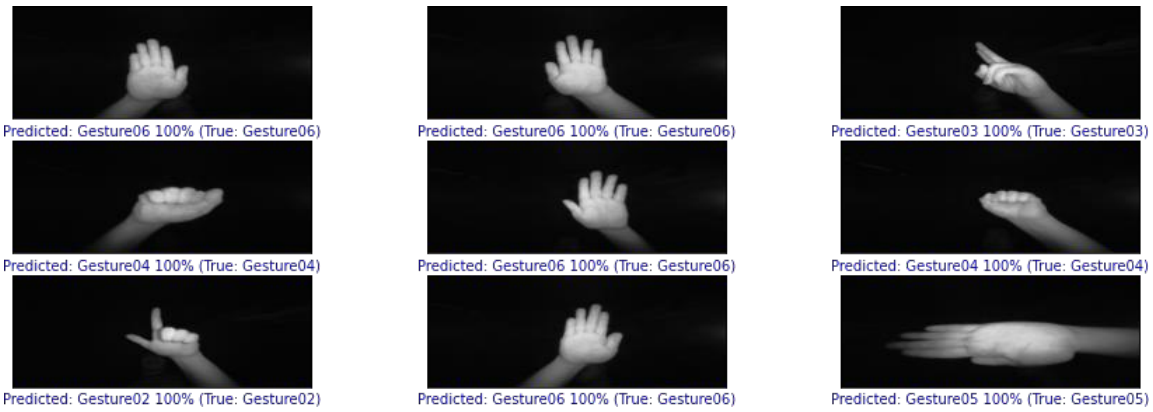

Figure 6. Image validation Result

	Predicted Gesture 01	Predicted Gesture 02	Predicted Gesture 3	Predicted Gesture 04	Predicted Gesture 05	Predicted Gesture 06
Actual Gesture 01	325	0	0	0	0	0
Actual Gesture 02	0	314	0	0	0	0
Actual Gesture 03	1	0	290	0	0	0
Actual Gesture 04	2	0	0	276	0	0
Actual Gesture 05	0	0	0	0	297	0
Actual Gesture 06	0	0	0	1	0	297

Figure 7. Matrix Confusion Table

```
57/57 [==============================] - 1s 12ms/step - loss: 0.0269 - accuracy: 0.9961
Test accuracy: 99.61%
```

Figure 8. Accuracy

4. CONCLUSION

Research using deep learning algorithms for hand gesture recognition has yielded promising results, indicating high accuracy in image classification tasks. Deep learning models, such as Convolutional Neural Networks (CNNs), have shown their ability to learn intricate patterns and features directly from raw image data. This enables them to accurately recognize and classify hand gestures. The practical implications of these findings are vast, ranging from improved human-computer interaction to advancements in sign language recognition and virtual reality. The accurate recognition of hand gestures can revolutionize user interfaces, making interactions with machines more intuitive and natural. Additionally, it can greatly benefit individuals with hearing impairments by enhancing communication and accessibility through sign language recognition systems. Ongoing research efforts aim to further optimize deep learning models and develop efficient architectures for real-time performance, while the availability of diverse and comprehensive training datasets remains crucial for achieving consistently high accuracy in hand gesture recognition tasks.

5. AUTHOR'S NOTE

The authors declare that there is no conflict of interest regarding the publication of this article. Authors confirmed that the paper was free of plagiarism.

DOI: https://doi.org/10.17509/seict.v2i1.34673
p- ISSN 2774-1656 e- ISSN 2774-1699

6. REFERENCES

Aamir, M., Rahman, Z., Abro, W. A., Tahir, M., and Ahmed, S. M. (2019). An optimized architecture of image classification using convolutional neural network. *International Journal of Image, Graphics and Signal Processing*, *10*(10), 30-39.

Abdulhai, B., Pringle, R., and Karakoulas, G. J. (2003). Reinforcement learning for true adaptive traffic signal control. *Journal of Transportation Engineering, 129*(3), 278-285.

Al-Doori, S. K. S., Taspinar, Y. S., and Koklu, M. (2021). Distracted driving detection with machine learning methods by CNN based feature extraction. *International Journal of Applied Mathematics Electronics and Computers*, *9*(4), 116-121.

Ando, R. K., Zhang, T., and Bartlett, P. (2005). A framework for learning predictive structures from multiple tasks and unlabeled data. *Journal of Machine Learning Research, 6*(11), 1817-1853.

Arifin, I., Haidi, R. F., and Dzalhaqi, M. (2021). Penerapan computer vision menggunakan metode deep learning pada perspektif generasi ulul albab. *Jurnal Teknologi Terpadu, 7*(2), 98-107.

Barbhuiya, A. A., Karsh, R. K., and Jain, R. (2021). CNN based feature extraction and classification for sign language. *Multimedia Tools and Applications, 80*(2), 3051-3069.

Choldun, M. I., and Surendro, K. (2018). Klasifikasi penelitian dalam deep learning. *Improve, 10*(1), 25-33.

Dasgupta, A., and Nath, A. (2016). Classification of machine learning algorithms. *International Journal of Innovative Research in Advanced Engineering (IJIRAE)*, *3*(3), 6-11.

Hu, F., Xia, G. S., Hu, J., and Zhang, L. (2015). Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sensing, 7*(11), 14680-14707.

Ju, C., Bibaut, A., and van der Laan, M. (2018). The relative performance of ensemble methods with deep convolutional neural networks for image classification. *Journal of Applied Statistics, 45*(15), 2800-2818.

Kaliyar, R. K., Goswami, A., and Narang, P. (2021). Fakebert: Fake news detection in social media with a bert-based deep learning approach. *Multimedia tools and applications, 80*(8), 11765-11788.

Khan, R. Z., and Ibraheem, N. A. (2012). Hand gesture recognition: A literature review. *International journal of artificial Intelligence and Applications, 3*(4), 161-174.

Kurniawan, A. A., and Mustikasari, M. (2021). Implementasi deep learning menggunakan metode CNN dan lSTM untuk menentukan berita palsu dalam bahasa Indonesia. *Jurnal Informatika Universitas Pamulang, 5*(4), 544-552.

Ridwang, R. (2017). Pengenalan Bahasa Isyarat Indonesia (SIBI) menggunakan leap motion controller dan algoritma data mining naïve bayes. *Jurnal Insypro (Information System and Processing), 2*(2), 1-8.

DOI: https://doi.org/10.17509/seict.v2i1.34673
p- ISSN 2774-1656 e- ISSN 2774-1699

Rijanandi, T., Rizaldy, A. A., Kridabayu, A. N., Devara, E. G. E., and Adhinata, F. D. (2022). Penerapan hair recognition menggunakan metode haar cascade classifier dan CNN deep learning. *Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer Universitas Al Asyariah Mandar*, *8*(1), 53-57.

Roihan, A., Sunarya, P. A., and Rafika, A. S. (2020). Pemanfaatan machine learning dalam berbagai bidang. *Indonesian Journal on Computer and Information Technology, 5*(1), 75-82.

Setiawan, D. (2018). Dampak perkembangan teknologi informasi dan komunikasi terhadap budaya. *Jurnal Simbolika: Research and Learning in Communication Study (E-Journal), 4*(1), 62-72.

Mesut Toğaçar, Burhan Ergen, and Zafer Cömert. (2019). BrainMRNet: brain tumor detection using magnetic resonance images with a novel convolutional neural network model. *Medical Hypotheses, 134*, 1-23.

Wahyuni, S., and Sulaeman, M. (2022). Penerapan algoritma deep learning untuk sistem absensi kehadiran deteksi wajah di PT Karya Komponen Presisi. *Jurnal Informatika Simantik, 7*(1), 12-21.

Yan, J., and Wang, X. (2022). Unsupervised and semi-supervised learning: the next frontier in machine learning for plant systems biology. *The Plant Journal, 111*(6), 1527-1538.

DOI: https://doi.org/10.17509/seict.v2i1.34673
p- ISSN 2774-1656 e- ISSN 2774-1699